AD-A069 455 CLEMSON UNIV S C DEPT OF MATHEMATICAL SCIENCES F/G 12/1

A BAYES SEQUENTIAL PROCEDURE FOR SELECTING THE MOST PROBABLE MU--ETC(U)

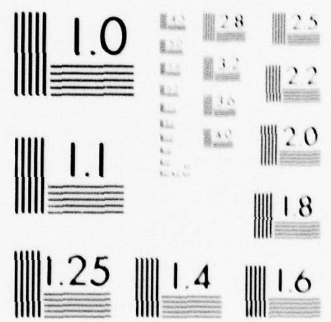
OCT 78 J T RAMEY, K ALAM N00014-75-C-0451

UNCLASSIFIED N100 NL

| OF |

AD
A069455

END
DATE
FILMED
7-79
DDC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

DEPARTMENT OF MATHEMATICAL SCIENCES

CLEMSON UNIVERSITY
Clemson, South Carolina

12 LEVEL II

12 16p.

6

A BAYES SEQUENTIAL PROCEDURE FOR SELECTING THE MOST PROBABLE MULTINOMIAL EVENT†.

BY

10 JAMES T. RAMEY, JR.[1]

[redacted]

KHURSHEED ALAM[2]

9 Technical rept.,

14

Report N10, TR-293

Technical Report #293

11 Oct[redacted]78

Research Supported in part by

THE OFFICE OF NAVAL RESEARCH

Task NR 042-271 Contract N00014-75-C-0451 15

[1]Department of Mathematics, Francis Marion College, Florence, S.C. 29501.

[2]Department of Mathematical Sciences, Clemson University, Clemson, S.C. 29631.

407 183

gw

A BAYES SEQUENTIAL PROCEDURE FOR SELECTING THE MOST PROBABLE MULTINOMIAL EVENT

By

James T. Ramey, Jr. and Khursheed Alam

Francis Marion College and Clemson University

ABSTRACT

This paper deals with a Bayes sequential sampling procedure for selecting the most probable event from a multinomial distribution whose parameters are distributed à priori according to a Dirichlet distribution. The given rule is compared with other sampling rules which have been considered in the literature.

Key words: Multinomial Distribution; Ranking and Selection; Bayes Sequential Decision Rules.

AMS Classification: 62F07; 62F15

This work was supported by the Office of Naval Research under Contract N00014-75-C-0451.

1. Introduction. Consider a multinomial population with $k \geq 2$ cells and the associated probability vector $\underset{\sim}{p} = (p_1,\ldots,p_k)$ where $\sum_{i=1}^{k} p_i = 1$. Let $p_{[1]} \leq p_{[2]} \leq \cdots \leq p_{[k]}$ denote the ordered values of $p_1,\ldots,p_k$. A number of statistical procedures based on sequential and non-sequential sampling rules have been considered in the literature for selecting the most probable event, that is, the event associated with $p_{[k]}$. The papers of Bechhofer et al. (1959), Cacoullos and Sobel (1966), Alam (1971) and Ramey and Alam (1979) may be cited for reference. Gupta and Nagel (1967) have considered a procedure for selecting a subset of the k events which includes the most probable event. In line with the classical approach of ranking and selection methods, the central theme of these papers is to find the least favorable configuration, that is, the value of $\underset{\sim}{p}$ for which the probability of a correct selection is minimized, and to determine the sample size for the fixed sample procedure and the expected sample size for the sequential sampling rule, for which the minimum probability of a correct selection is equal to a given number p^*. In this paper we consider the selection problem from a Bayesian approach.

The multinomial probability function is given by

$$(1.1) \qquad f(x_1,\ldots,x_n) = \binom{n}{x_i,\ldots,x_k} p_1^{x_1} \cdots p_k^{x_k}$$

where $\sum_{i=1}^{n} x_i = n$. Let $D(\nu_1,\ldots,\nu_k)$ denote the Dirichlet distribution, given by the density function

$$(1.2) \qquad g(\underset{\sim}{p}) = \frac{\Gamma(\nu)}{\Gamma(\nu_1) \ldots \Gamma(\nu_k)} \, p_1^{\nu_1 - 1} \ldots p_k^{\nu_k - 1}$$

$$\nu = \textstyle\sum_{i=1}^{k} \nu_i, \; \nu_i > 0, \; i = 1, \ldots, k.$$

From (1.1) and (1.2) it is seen that the Dirichlet distribution is a conjugate prior distribution for $\underset{\sim}{p}$. We consider a Bayes sequential sampling procedure for selecting the most probable event, assuming a Dirichlet prior distribution for $\underset{\sim}{p}$ and the loss function being given as follows: Let $c > 0$ denote the cost of sampling per unit observation and let d_i denote the decision to select the event associated with the ith cell as the most probable event, after the sampling is stopped. If n observations have been taken the loss is given by

$$(1.3) \qquad L(d_i, \underset{\sim}{p}) = nc + (p_{[k]} - p_i) .$$

The theory of the Bayes sequential sampling procedure is given in Section 2 and its application is shown in Section 3. In Section 4 we compare its risk, that is the expected loss, with the risk of the fixed sample procedure of Bechhofer and the sequential sampling rule of Cacoullos and Sobel.

2. <u>Bayes sequential sampling rule</u>. Suppose that $\underset{\sim}{p}$ is distributed a priori according to the Dirichlet distribution $D(\nu_1, \ldots, \nu_k)$. Then the posterior distribution of $\underset{\sim}{p}$ given the observation $\underset{\sim}{x} = (x_1, \ldots, x_k)$ is $D(\nu_1 + x_1, \ldots, \nu_k + x_k)$. Marginally, p_i is distributed according to $D(\nu_i, \nu - \nu_i)$ and its mean is given by

$$E\, p_i = \nu_i/\nu \ . \tag{2.1}$$

After the sampling is stopped, the terminal decision should be a Bayes rule with respect to the posterior distribution of $\underset{\sim}{p}$. It follows from (1.3) and (2.1) that the terminal decision is d_i if

$$x_i + \nu_i = \max\,(x_1+\nu_1,\ldots,x_k+\nu_k). \tag{2.2}$$

From (1.3) it is seen that $n < 1/c$ since the loss from an immediate decision without taking an observation is < 1. Therefore, we need consider only bounded sequential procedures. Let $D^* = D(\nu_1,\ldots,\nu_k)$ and let $\rho_j(D^*)$ denote the Bayes risk due to an optimal sequential sampling rule in which not more than j observations should be taken. From (1.3) and (2.1) it is seen that

$$\rho_o(D^*) = E\, p_{[k]} - \max\,(\nu_1,\ldots,\nu_k)/\nu. \tag{2.3}$$

The marginal distribution of a single observation $\underset{\sim}{y} = (y_1,\ldots,y_k)$ is given by

$$P\{y_i = 1,\ y_j = 0,\ j \neq i\} = \frac{\nu_i}{\nu},\ i = 1,\ldots,k.$$

Let $D^*(\underset{\sim}{y})$ denote the posterior distribution, given $\underset{\sim}{y}$. From backward induction (see e.g. Degroot (1970), §12.5), a recursive formula for the Bayes risk is given by

$$(2.4)\qquad \rho_{j+1}(D^*) = \min\{\rho_0(D^*),\ E\,\rho_j(D^*(\underset{\sim}{y}) + c\}$$

$$= \min\{\rho_0(D^*),\ \textstyle\sum_{i=1}^{k} \frac{\nu_i}{\nu}\,\rho_j(D_i^*) + c\},$$

$$j = 0,1,\ldots,m-1$$

where m denotes the largest integer $\leq 1/c$ and $D_i^* =$ $D(\nu_1,\ldots,\nu_{i-1},\nu_i+1,\nu_{i+1},\ldots,\nu_k)$. Putting $j = 0$ in (2.4) we get

$$(2.5)\qquad \rho_1(D^*) = \min\{\rho_0(D^*),\ E\,\rho_0(D^*(\underset{\sim}{y})) + c\}$$

$$= \min\{\rho_0(D^*),\ E\,p_{[k]} + c - \textstyle\sum_{i=1}^{k} \frac{\nu_i}{\nu(\nu+1)}$$

$$\max(\nu_1,\ldots,\nu_{i-1},\nu_i+1,\nu_{i+1},\ldots,\nu_k)$$

$$= \rho_o(D^*) + \min\{0,\ c + \frac{\max(\nu_1,\ldots,\nu_k)}{\nu} - \textstyle\sum_{i=1}^{k} \frac{\nu_i}{\nu(\nu+1)}$$

$$\cdot\ \max(\nu_1,\ldots,\nu_{i-1},\nu_i+1,\nu_{i+1},\ldots,\nu_k)\ .$$

The recursive relation (2.4) gives the stopping rule for the Bayes sequential sampling procedure, as follows: Take no observation if $\rho_0(D^*) \leq \rho_m(D^*)$. Otherwise, take observations sequentially. Stop sampling after taking n observations if $\rho_0(D^*(\underset{\sim}{x})) \leq \rho_{m-n}(D^*(\underset{\sim}{x}))$.

3. <u>Application</u>. Consider the special case in which $\nu_1 = \nu_2 = \ldots = \nu_k = \nu_0$, say. Let $D_0 = D(\nu_0,\ldots,\nu_0)$ and let p_0 denote

the expected value of $p_{[k]}$ with respect to the prior distribution D_0. From (2.3) and (2.5) we get

$$\rho_0(D_0) = p_0 - \frac{1}{k},$$

$$\rho_1(D_0) = p_0 - \max\left(\frac{1}{k}, \frac{\nu_0+1}{k\nu_0+1} - c\right).$$

Table I below shows the continuation set, that is, the set of sample points or the cell counts where the sampling should be continued, for $\nu_0 = 1$, $c = .01$ and $k = 2(1)5$. The complement of the continuation set represents the set of sample points where the sampling should be stopped. The value of n in the table represents the number of observations or the stage of sampling. For example, let $k = 2$ and $n = 10$. The table shows that the sample point (5.5) lies in the continuation set. Therefore, if 10 observations have been taken and the cell counts are (5,5) then at least one more observation should be taken. On the other hand, if the cell counts are (6,4), then the sampling is stopped and the cell associated with the count number 6 should be selected as the most probable event.

4. Comparison of three sampling rules. The sampling scheme of Cacollous and Sobel (1966) in which observations are taken sequentially until the largest count in any cell is equal to a given positive integer N, has been called inverse sampling (IS). It would be interesting to compare the Bayes sequential

sampling rule given above with the IS and the fixed sample (FS) rule of Bechhofer (1959). Suppose that $\underset{\sim}{p}$ is distributed a priori according to the Dirichlet distribution D_0. Since the terminal decision for both the IS and FS rules is to select the event associated with the largest cell count, it is Bayes with respect to D_0.

The value of $\rho_m(D_0)$, the risk of the Bayes sequential sampling rule is obtained from (2.4). Formulas for the risk of FS and IS being denoted by $\rho(FS)$ and $\rho(IS)$, respectively, are given in the Appendix. Let $\rho^*(FS)$ denote the minimum value of $\rho(FS)$ minimized for all values of n, the fixed sample size. Similarly, let $\rho^*(IS)$ denote the minimum value of $\rho(IS)$, minimized for all values of N. Table II below gives values of $\rho(FS)$ and $\rho(IS)$ for certain values of ν_0, c and k, where

$$\rho(FS) = \frac{\rho^*(FS) - \rho_m(D_0)}{\rho^*(FS)} \times 100, \quad \rho(IS) = \frac{\rho^*(IS) - \rho_m(D_0)}{\rho^*(IS)} \times 100.$$

The values of $\rho(FS)$ and $\rho(IS)$ represent the percentage reduction in the risk due to the optimal Bayes sequential sampling rule, compared with the FS and IS sampling rules, respectively. It is seen from the table that $\rho(FS) > \rho(IS)$ for any given value of (c, ν_0, k), and that $\rho(FS)$ and $\rho(IS)$ are both decreasing in c for any given value of (ν_0, k).

APPENDIX

5. Values of $\rho_m(D_0)$, $\rho(FS)$ and $\rho(IS)$. From (1.1) and (1.2) we obtain the marginal distribution of $\underset{\sim}{x}$, given by the probability function

$$(5.1)\qquad h(\underset{\sim}{x}) = \frac{\Gamma(k\nu_0)}{\Gamma(k\nu_0+n)(\Gamma(\nu_0))^k} \binom{n}{x_1,\dots,x_k} \Pi_{i=1}^{k} \Gamma(\nu_0+x_i) .$$

The risk of FS is given by

$$(5.2)\qquad \rho(FS) = p_0 + nc - (k\nu_0+n)^{-1}(\nu_0 + \sum_{x_1+\dots+x_k = n} \max(x_1,\dots,x_k)h(\underset{\sim}{x}))$$

where n denotes the fixed sample size and $p_0 = E\, p_{[k]}$, the expected value of $p_{[k]}$ with respect to the prior distribution D_0. Similarly, the risk of IS is given by

$$(5.3)\qquad \rho(IS) = p_0 - (\nu_0+N-1) \sum_{n = N-1}^{k(N-1)} \sum_{\substack{x_1+\dots+x_k = n+1 \\ x_{[k-1]} < x_{[k]} = N}} (k\nu_0+n)^{-1} \left(\frac{\nu_0+N}{k\nu_0+n+1} - (n+1)c\right) h(\underset{\sim}{x}^*)$$

where $x_{[i]}$ denotes the ith smallest amongst $x_1,\dots,x_k$, and

$$x_i^* = \begin{cases} x_i & \text{if } x_i \neq N \\ x_i - 1 & \text{if } x_i = N . \end{cases}$$

The value of $\rho_m(D_0)$ is obtained from (2.4).

It is seen that the formulas for the three risk functions, given above, involve the computation of $E\, p_{[k]}$, the expected value of p_k with respect to a Dirichlet prior distribution. The value of $E\, p_{[k]}$ is derived as follows. Let $y_1,\ldots,y_k$ be k independent random variables and let y_i be distributed according to the gamma distribution G_{ν_i} with ν_i degrees of freedom, $i = 1,\ldots,k$. Let

$$z_i = y_i/(\textstyle\sum_{i=1}^{k} y_i).$$

Then $z = (z_1,\ldots,z_k)$ is distributed according to the Dirichlet distribution $D(\nu_1,\ldots,\nu_k)$. From the scale invariant property of the statistic $\max\,(z_1,\ldots,z_k)$ it follows that $\sum_{i=1}^{k} y_i$ and $\max\,(z_1,\ldots,z_k)$ are independently distributed. Hence

$$\begin{aligned} E\, p_{[k]} &= E \max\,(z_1,\ldots,z_k) \\ &= \frac{E \max\,(y_1,\ldots,y_k)}{E \sum_{i=1}^{k} y_i} \\ &= (\nu_1+\ldots+\nu_k)^{-1} E \max\,(y_1,\ldots,y_k) \\ &= \nu^{-1} \int_0^\infty (1-\Pi_{i=1}^{k} G_{\nu_i}(x))\,dx. \end{aligned} \tag{5.3}$$

From (5.3) we have

$$p_0 = (k\nu_0)^{-1} \int_0^\infty (1-G_{\nu_0}^{k}(x))\,dx.$$

Thus, $k\nu_0\rho_0$ represents the expected value of the largest order statistic in a sample size k from the gamma distribution G_{ν_0}. The moments of order statistics from the gamma distribution have been tabulated by Harter (1969).

TABLE I - Continuation set of sample points for $\nu_0 = 1$, $c = .01$

n	k = 2	k = 3	k = 4	k = 5
1	(1,0)	(1,0,0)	(1,0,0,0)	(1,0,0,0,0)
2	(1,1)	(1,1,0)	(1,1,0,0)	(1,1,0,0,0)
3	(2,1)	(1,1,1), (2,1,0)	(1,1,1,0), (2,1,0,0)	(1,1,1,0,0), (2,1,0,0,0)
4	(2,2)	(2,1,1), (2,2,0)	(1,1,1,1), (2,1,1,0) (2,2,0,0)	(1,1,1,1,0), (2,1,1,0,0), (2,2,0,0,0)
5	(3,2)	(2,2,1), (3,2,0)	(2,1,1,1), (2,2,1,0) (3,2,0,0)	(1,1,1,1,1), (2,1,1,1,0), (2,2,1,0,0)
6	(3,3)	(2,2,2), (3,2,1) (3,3,0)	(2,2,1,1), (2,2,2,0) (3,3,0,0)	(2,1,1,1,1), (2,2,1,1,0), (2,2,2,0,0)
7	(4,3)	(3,2,2), (3,3,1) (4,3,0)	(2,2,2,1), (3,2,2,0) (3,3,1,0)	(2,2,1,1,1), (2,2,2,1,0), (3,2,2,0,0)
8	(4,4)	(3,3,2), (4,4,0)	(2,2,2,2), (3,2,2,1) (3,3,1,1), (3,3,2,0)	(2,2,2,1,1), (2,2,2,2,0), (3,3,2,0,0)
9	(5,4)	(3,3,3), (4,4,1) (5,4,0)	(3,2,2,2), (3,3,2,1) (3,3,3,0)	(2,2,2,2,1), (3,2,2,2,0), (3,3,2,1,0) (3,3,3,0,0)
10	(5,5)	(4,3,3), (4,4,2) (5,5,0)	(3,3,2,2), (3,3,3,1) (4,3,3,0)	(2,2,2,2,2), (3,3,2,1,1), (3,3,2,2,0) (3,3,3,1,0)
11	(6,5)	(4,4,3), (5,5,1)	(3,3,3,2), (4,4,3,0)	(3,2,2,2,2), (3,3,2,2,1), (3,3,3,1,1) (3,3,3,2,0)
12	(6,6)	(4,4,4), (5,5,2)	(3,3,3,3), (4,4,3,1) (4,4,4,0)	(3,3,2,2,2), (3,3,3,2,1), (3,3,3,3,0)
13	ϕ	(5,4,4), (5,5,3)	(4,3,3,3), (4,4,3,2) (4,4,4,1)	(3,3,3,2,2), (3,3,3,3,1)
14		(5,5,4)	(4,4,3,3), (4,4,4,2)	(3,3,3,3,2)
15		(5,5,5)	(4,4,4,3)	(3,3,3,3,3)
16		ϕ	(4,4,4,4)	ϕ
17			ϕ	

ϕ - denotes the null set

TABLE II - Values of ρ(FS) and ρ(IS)

k = 2

ν_0	.5			1			2			3		
c	.01	.005	.001	.01	.005	.001	.01	.005	.001	.01	.005	.001
ρ(FS)	19.9	25.4	39.2	16.5	21.7	33.7	12.4	17.4	28.5	9.7	14.4	27.9
ρ(IS)	9.8	12.5	25.8	8.1	10.6	21.8	6.1	8.1	18.5	3.6	6.6	16.5

k = 3

ρ(FS)	21.6	24.4	33.8	18.4	19.8	28.2	13.3	16.0	23.2	8.8	14.0	20.5
ρ(IS)	6.7	9.2	19.8	4.7	6.8	16.0	2.7	5.1	12.7	0.7	4.5	11.1

k = 4

ρ(FS)	19.9	21.7	30.2	15.7	18.1	27.0	11.5	14.2	20.0	4.7	12.4	17.6
ρ(IS)	4.7	7.0	16.4	3.3	5.1	12.9	0.7	3.7	9.9	0.0	2.1	8.4

k = 5

ρ(FS)	18.1	20.4	24.8	13.4	16.7	22.0	8.0	13.3	17.5	1.0	10.4	15.4
ρ(IS)	3.6	5.7	11.2	2.0	5.2	10.5	.0	1.8	7.8	0.0	1.6	6.4

REFERENCES

[1] Alam, K. (1971). On selecting the most probable category. Technometrics (13), pp. 843-850.

[2] Bechhofer, R., Elmaghrabi, S. and Morse, N. (1959). A single-sample multiple decision procedure for selecting the multinomial event which has the highest probability. Ann. Math. Statist. (30) pp. 102-119.

[3] Cacoullos, T. and Sobel, M. (1966). An inverse sampling procedure for selecting the most probable event in a multinomial distribution. Proceedings of the International Symposium on Multivariate Analysis, Ed. Krishnaiah, P.R., pp. 423-455, Academic Press, Inc., New York.

[4] Gupta, S. S. and Nagel, K. (1967). On selection and ranking procedures and order statistics from the multinomial distribution. Sankhya, Series B 29, pp. 1-17.

[5] Harter, H. L. (1969). Order statistics and their use in testing and estimation. Volume 2. Aerospace Research Laboratories Report.

[6] Ramey, Jr., J. T. and Alam, K. (1979). A sequential procedure for selecting the most probable multinomial event. Biometrika (to appear).

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER N100	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A Bayes Sequential Procedure For Selecting The Most Probable Multinomial Event		5. TYPE OF REPORT & PERIOD COVERED Technical 6. PERFORMING ORG. REPORT NUMBER TR #293
7. AUTHOR(s) James T. Ramey Khursheed Alam		8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0451
9. PERFORMING ORGANIZATION NAME AND ADDRESS Clemson University Dept. of Mathematical Sciences Clemson, South Carolina 29631		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 042-271
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Code 436 Arlington, Va. 22217		12. REPORT DATE October, 1978 13. NUMBER OF PAGES 12
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Multinomial Distribution; Ranking and Selection; Bayes Sequential Decision Rules.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This paper deals with a Bayes sequential sampling procedure for selecting the most probable event from a multinomial distribution whose parameters are distributed à priori according to a Dirichlet distribution. The given rule is compared with other sampling rules which have been considered in the literature.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)